



Merlo, J., Wagner, P., & Leckie, G. (2019). A simple multilevel approach for analysing geographical inequalities in public health reports: The case of municipality differences in obesity. *Health and Place*, 58, [102145]. <https://doi.org/10.1016/j.healthplace.2019.102145>

Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
[10.1016/j.healthplace.2019.102145](https://doi.org/10.1016/j.healthplace.2019.102145)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the final published version of the article (version of record). It first appeared online via Elsevier at <https://www.sciencedirect.com/science/article/pii/S1353829218312371> . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>



A simple multilevel approach for analysing geographical inequalities in public health reports: The case of municipality differences in obesity

Juan Merlo^{a,*}, Philippe Wagner^{a,c}, George Leckie^{a,b}

^a Research Unit of Social Epidemiology, Faculty of Medicine, University of Lund, Jan Waldenströms Street 35, SE-20502 Malmö, Sweden

^b Centre for Multilevel Modelling, University of Bristol, 35 Berkeley Square, Bristol, BS8 1JA, UK

^c Centre for Clinical Research Västmanland, Uppsala University, Västmanland Hospital, 721 89 Västerås, Sweden

ARTICLE INFO

Keywords:

Public health
Geographical differences
Multilevel analysis
Discriminatory accuracy
Obesity

ABSTRACT

The epidemiological analysis of geographical inequalities in individual outcomes is a fundamental theme in public health research. However, many traditional studies focus on analysing area differences in averages outcomes, disregarding individual variation around such averages. In doing so, these studies may produce misleading information and lead researchers to draw incorrect conclusions. Analysing individual and municipality differences in body mass index (BMI) and overweight/obesity status, we apply an analytical approach based on the multilevel analysis of individual heterogeneity and discriminatory accuracy (MAIHDA). This analytical approach may be viewed as a reorganization of existing multilevel modelling concepts in order to provide a systematic approach to simultaneously considering both differences between area averages and individual heterogeneity around those averages. In doing so, MAIHDA provides an improved approach to the quantification and understanding of geographical inequalities as compared with traditional approaches.

1. Introduction

In Sweden, as in many countries, equity is a fundamental goal of the healthcare system and there is a clear political determination to reduce health inequalities. An important step to achieving this goal is the epidemiological analysis of disparities in health outcomes and healthcare utilization between societal groups and across different geographical areas (The Commission for Health on Equal Terms, 2017). Traditional epidemiological studies report area differences in average health outcomes such as disease prevalence. These studies frequently report the existence of “substantial” and “significant” geographical variation. They also stimulate societal interest in public health questions as well as development work to reduce health inequalities. However, a limitation of these studies is that they typically disregard the variation in individual outcomes around population group and area averages (Merlo, 2003). In doing so, these studies run the risk of inefficiently allocating resources and inaccurately labelling groups of individuals or areas with relatively “bad” average outcomes. As Hans Rosling pointed out (Rosling et al., 2018) (page 61) “If you were able to investigate the individual variation around the averages, you would probably see the groups overlap and there's probably no gap at all” (Author's translation from Swedish). However, there is no hesitation that “places” and social contexts are fundamental for understanding

individual health disparities. Rosling's observation depends tremendously upon the health outcome in question and neither contradicts the rest of his work nor the ideas of Durkheim (1964) and Rose (1992) on the existence of emerging social facts and sick populations. However, the accuracy of traditional epidemiological methods, based on comparing averages, in identifying those places and contexts that really matter for specific individual outcomes is still insufficiently known. We have long expressed related ideas in earlier publications (Merlo et al., 2004; Merlo, 2003; Merlo et al., 2009; Merlo et al., 2012; Merlo and Mulinari, 2015; Merlo et al., 2017).

In this article, we present a framework for the analysis of geographical inequalities in health, based on the *multilevel analysis of individual heterogeneity and discriminatory accuracy* (MAIHDA). MAIHDA is not a new methodology but it may be viewed as a reorganization of existing multilevel modelling concepts in order to systematically and simultaneously considers group differences in average outcomes and the extent of individual variation around such averages (Merlo, 2014; Merlo, 2018a). From this perspective, geographical differences are not assessed by simple measures of spatial or small area variance but as the share of the total individual variance that are at the geographical (i.e., area) level. The approach quantifies the accuracy of the information provided by the area differences to classify persons according to the outcome. MAIHDA can help us distinguish between situations where

* Corresponding author.

E-mail addresses: juan.merlo@med.lu.se (J. Merlo), phille.wagner@gmail.com (P. Wagner).

<https://doi.org/10.1016/j.healthplace.2019.102145>

Received 9 December 2018; Received in revised form 1 May 2019; Accepted 27 May 2019

1353-8292/ © 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

there are large geographic differences and situations where there are not.

Scholars applying/developing multilevel modelling will recognize the concepts used in our study. For other readers, however, several features of the multilevel analysis will likely prove innovative. In addition, researchers working in the field of small area variation and spatial analyses might also initially find challenging the use of MAIHDA, as the spatial approach exclusively focuses on the spatial component of total individual variation (Merlo et al., 2012).

Using non-technical language and avoiding equations, we aim to make our arguments accessible to as broad an audience as possible. Performing a simple multilevel analysis of individual and municipality differences in body mass index (BMI) and overweight/obesity status, we illustrate why traditional analyses may lead researchers to draw misleading conclusions when they ignore individual variation around area averages. We argue that an analytical strategy based on MAIHDA, offers considerable benefits for the study of health inequalities across areas and population groups.

The disposition of our article is as follows. We start by a short description of the population and variables used for the empirical examples. Thereafter we introduce the multilevel modelling methodology and explain some fundamental concepts. Finally, we use those concepts within a systematic MAIHDA framework for quantifying geographical inequalities in health. The systematic framework and the empirical examples we describe are the simplest. More elaborated theory and examples that include adjustment for individual variables and the analysis of contextual variables can be read elsewhere. See for instance (Merlo et al., 2016).

2. Population and methods

2.1. Study sample

We use data from the Health Survey for Skåne 2008. We study 28,198 participants aged between 18 and 80 years residing in the 33 municipalities of Skåne, the southernmost county of Sweden. We exclude 893 individuals with missing values on BMI and then a further 83 individuals with outlier values on BMI. The final study sample consists of 27,222 individuals. See elsewhere for details of the survey (Rosvall et al., 2009).

2.2. Assessment of variables and public health targets

We analysed self-reported BMI (Nyholm et al., 2007) both as a continuous variable, measured in kg/m², and as a binary variable “overweight/obese”, defined as a BMI ≥ 25.0 kg/m².

2.3. Multilevel modelling

We fit multilevel models as the data have two levels of analysis with individuals (level 1) nested within municipalities (level 2). We explain the intuition for all our analyses, but we omit more advanced methodological and statistical details in the interests of maintaining accessibility and because the reader can readily find such information in the literature (Centre for Multilevel Modelling, <http://www.bristol.ac.uk/cmm/learning/online-course/>, Merlo et al., 2006; Merlo et al., 2005a,b,c; Duncan et al., 1998; Snijders and Bosker, 1999; Subramanian et al., 2003; Merlo et al., 2012).

2.3.1. Multilevel linear regression and the concept of the variance partition coefficient (VPC)

In our first empirical example, we analyse the individual-level continuous BMI variable outcome using a two-level linear regression model. This example facilitates understanding of some of the key concepts in multilevel analysis, including: the total variance in individual outcomes, the variance in municipality averages around the overall

county average, the variance in individual outcomes around municipality averages, and the concept of the variance partition coefficient (VPC). The VPC measures the percentage of total variance in an individual health indicator found at the municipality level (Goldstein et al., 2002; Merlo et al., 2006). The total individual variance is the sum of the variance in health outcomes between the municipalities (V_M) and the variance between individuals within the municipalities (V_I).

$$\text{VPC} \equiv \text{ICC} = 100 \times \frac{V_M}{V_M + V_I}$$

The theoretical range of the VPC is from 0% to 100%. A large VPC therefore indicates the existence of substantial geographic disparities, while a small VPC indicates that any geographic inequities are dwarfed by the individual variation within each area. In two-level settings such as this, the VPC corresponds with the intraclass correlation coefficient (ICC). The ICC measures the expected correlation in BMI between two individuals from the same municipality.

To understand better the meaning of the VPC and ICC, we present two extreme situations. In the first situation, we assume the average outcome value is the same in every area or it only differs because of statistical uncertainty. This would also be the case if the “areas” were simply groups of individuals randomly sampled from the population rather than groups of individuals systematically sampled from different contexts (see (Snijders and Bosker, 1999), Example 3.1, p.18). In this case, the VPC or ICC would be equal to 0%, and there would therefore be no area contextual effects on individual outcome (i.e., BMI). In this situation, individual outcomes would be independent within areas; the “sample” would not be correlated at all. In the second situation, we assume all individual in an area have the same outcome value, which in turn differs from the common outcome values seen in other areas. In this case, the VPC or ICC would be to 100%. Individual outcomes would be perfectly correlated within areas.

When studying repeated measurements within individuals (i.e., in this case the measurements are at level 1 and individuals are at level 2) the VPC will often be very high. This is because the “individual” is a well-defined system/context and the measurements are highly correlated over time for obvious reasons. However, in the case of individuals within areas, the context that influences the individuals is not necessarily captured by the modifiable area unit definition used by the researcher (Merlo et al., 2009). A geographical unit definition maybe relevant for some individual outcomes but not for others, but in this last case it does not mean the social or geographical context is always irrelevant. Rather, it suggests the relevant context for that individual outcome has not been identified. For instance, in a previous study we found that the neighbourhoods in Malmö were very relevant for understanding an individual's use of private general practitioners (VPC = 56%), but the same neighbourhoods were of little relevance for predicting psychotropic drug use (VPC = 1%) (Merlo et al., 2016). That is, we found pronounced geographical disparities in the use of private general practitioners, but effectively no geographic disparities in relation to the use of psychoactive drugs. See elsewhere for a longer discussion on these issues (Merlo et al., 2012).

A fundamental idea in our approach is that we do not consider individual and geographical disparities as separated phenomena of interest. Rather, we study them simultaneously. From this perspective, the total individual disparities may be very relevant but their geographical component small or large. Thus, the VPC constitutes a far more informative measure for evaluating geographical differences than traditional measures based only on differences between area averages (Merlo, 2014).

2.3.1.1. A visual explanation of the VPC concept. To facilitate the visual understanding of this idea, Fig. 1 presents the distribution of individual BMI in two hypothetical municipalities. The difference in average individual health outcomes between the two municipalities in scenario A (D_A) is as large as that presented in scenario B (D_B), but

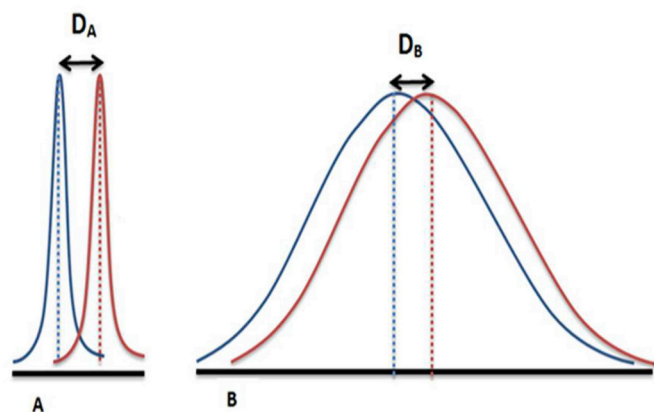


Fig. 1. The figure represents the individual distributions of a continuous variable, for instance body mass index (BMI), in two municipalities. The difference in average health outcomes between the two municipalities in scenario A (D_A) is as large as that presented in scenario B (D_B), but the individual variation around the average values is much smaller in scenario A than in scenario B. Clearly, the same difference in the average value between two municipalities is possible with very different degrees of individual variation within those municipalities.

the individual variation around the average values is much smaller in scenario A than in scenario B. Clearly, the same difference in the average value between two municipalities is possible with very different degrees of individual variation within those municipalities.

In the first scenario (Fig. 1A) the individual variation in BMI around the average values is very small in relation to the difference between the municipality average values. Since there is almost no overlap between the two municipalities' individual distributions, we can say that a large share of the total variation in individual health outcomes operates at the municipality level. Thus, when the overlap is very small (i.e., a high VPC), we can initially say that the difference between the two municipalities is very large.

In contrast, in the second scenario, (Fig. 1B), the variation in individual outcomes around the area averages is very large in relation to the difference between the municipality average values. In this scenario, there is substantial overlap across the two distributions (i.e., a low VPC) and, therefore, the difference in average outcomes between the two municipalities is very small.

2.3.2. Multilevel logistic regression and the concept of area under the receiver operating characteristic curve (AUC) statistic

In our second empirical example, we apply multilevel logistic regression to analyse individuals' overweight/obese status. Here the same

general multilevel concepts and interpretations apply as in multilevel linear regression although estimation and interpretation become somewhat more complex. One potential difficulty for some readers is that the VPC in multilevel logistic regression is expressed not in terms of the binary health outcome of interest, but in terms of a latent continuous propensity of being overweight/obese. Thus, an increasingly popular alternative to the VPC when performing multilevel logistic regression is to report the area under the receiver operating characteristic curve (AUC) statistic. The AUC is widely used within medicine and public health for evaluating the quality of diagnostic and screening tests as well as risk factors (Pepe et al., 2004). In the present example, the AUC is based on the predicted probability of the individual overweight/obese outcome as a function of the municipality random effects from the two-level logistic regression model. See elsewhere for detailed explanations (Merlo et al., 2016).

The AUC is constructed by plotting the true positive fraction (TPF) (i.e., sensitivity; the proportion of actual overweight/obese individuals that are correctly identified as such) against the false positive fraction (FPF) (i.e., $1 - \text{specificity}$; the proportion of actual normal weight individuals that are wrongly identified as overweight/obese) for different binary classification thresholds of the predicted probabilities. The AUC measures the accuracy of knowing where the individual resides (i.e., the municipalities) for discriminating overweight/obese individuals from individuals of normal weight. Formally, the AUC can be defined as the probability that a randomly selected overweight/obese individual will have a higher predicted probability than a randomly selected normal weight individual. In our simple example, the predicted probability is only dependent on municipality of residence and so the AUC is simply the probability that a randomly selected overweight/obese individual resides in a municipality with a higher prevalence of overweight/obese individuals than does a randomly selected normal weight individual. The AUC takes a value between 0.5 and 1.0 where 0.5 corresponds to the municipalities having no discriminatory accuracy and 1.0 corresponds to perfect discrimination.

Fig. 2 presents the relationship between the AUC and the VPC. We produced this relationship via simulation (we provide here a technical explanation of the procedure but the reader may disregard it without losing continuity). Specifically, we fitted two-level variance-components logistic regression models to multiple simulated datasets each with 100 neighbourhoods and 100 individuals per neighbourhood. We then predicted the individual probabilities of a positive outcome and using those, we calculated the AUC statistic. In each simulation, we held the population-averaged prevalence constant at 50%, which is close to the population average prevalence of obesity/overweight in our sample. We varied the VPCs across the simulations from 0 to 100% in increments of 1. For each value of the VPC, we repeated the simulations 1000 times and averaged the resulting AUC statistics to obtain the smoothed relationship plotted in Fig. 2.

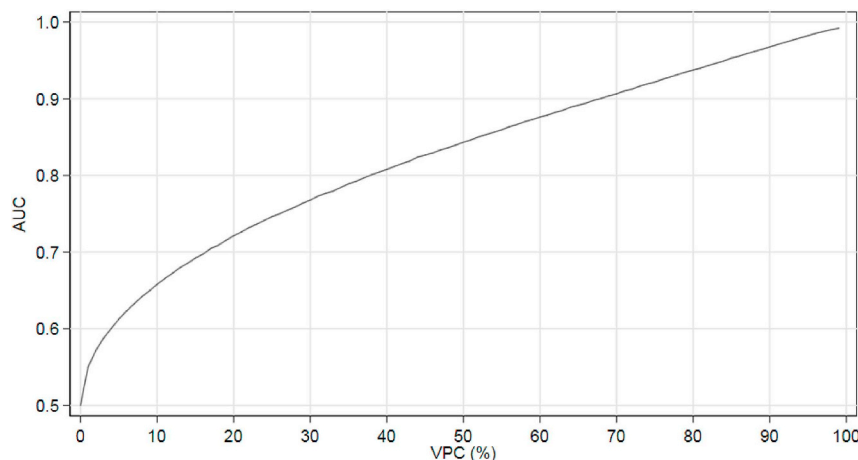


Fig. 2. The figure illustrates the relationship between the variance partition coefficient (VPC) and the area under receiver operating characteristics curve (AUC) for a binary individual health outcome. The figures are based on simulated data. Specifically, balanced two-level datasets (100 areas with 100 individuals per area) where the population average prevalence is 50% and where we vary the VPC from 0 to 100% in increments of 1. For each simulated dataset, we calculate the AUC based on the individual predicted probabilities. We repeat the simulations 1000 times to average away the simulation variability.

For example, a VPC of 5% corresponds to an AUC of 0.61. Put into words, in a scenario where we observe that 5% of the total variation in an individual health outcome lies between areas, we would observe a 61% probability that a randomly selected overweight/obese individual will reside in a municipality with a higher prevalence of overweight/obese individuals than does a randomly selected normal weight individual. When the VPC equals 0% the AUC equals 50%, while when the VPC equals 100% the AUC also equals 100%.

3. Proposal of a simple MAIHDA framework for quantifying geographical inequalities in health

The underlying reason for doing our analysis is to monitor the possible existence of geographical inequalities in health as there is a clear political determination to reduce those inequalities. We may analyse municipalities, as in our empirical examples, because the municipal context may influence the BMI of the individuals and also the municipality is an important public health arena for intervention. We denominate our analytical framework as MAIHDA in order to stress that we focus on the analysis of both averages and the individual heterogeneity around the averages. This framework includes four steps that need be considered to achieve a complete analysis of geographical inequalities in health. However, more elaborated strategies are of course also possible.

1. - Identification of a health target or benchmark average value.
2. - Describing and visualizing geographical differences.
3. - Quantifying the size of the geographical differences.
4. - Interpretation of the results.

3.1. Identification of a health target or benchmark average value

Reducing health inequalities is not enough by itself; we also need a target expressing a desirable average level of health in the population. For the purposes of illustrating our arguments, we assume that the overall public health target in Skåne and its municipalities was to achieve average BMI below 25.0 kg/m² (or, in the case of analysing the binary overweight/obese measure, to achieve a prevalence of overweight/obese individuals below 50%).

3.2. Describing and visualizing geographical differences

In this step we perform a standard analysis of the municipality differences in relation to the benchmark. In addition, we carry out an appropriate visual assessment of the geographical differences. For this purpose, we can use for instance, health league tables, funnel plots, choropleth maps or atlases of geographical variation to compare area averages. However, for a correct interpretation that effectively communicate the right information we need to satisfy the following conditions.

Table 1

This table proposes a framework for performing geographical comparisons of a specific public health indicator. For this purpose, we need information about the *target indicator value*. That is, whether the average health indicator (e.g., the average BMI value or the prevalence of overweight/obesity) has “insufficiently”, “closely” or “fully” reached a predetermined target *benchmark* level. We also need information on the *size of the observed geographical differences* as assessed by the variance partition coefficient (VPC) or the area under receiver operating characteristics curve (AUC). Combining this information, we obtain 18 different scenarios useful for the evaluation.

Absence or small geographical differences	VPC (%)	AUC	Target indicator value reached		
			Insufficiently	Closely	Fully
● Absent	0 to 1	0.50 to 0.55	A	B	C
● Very small	1 to 5	0.55 to 0.61	D	E	F
● Small	5 to 10	0.61 to 0.66	G	H	I
Large geographical differences					
● Less large	10 to 20	0.66 to 0.72	J	K	L
● Fairly large	20 to 30	0.72 to 0.77	M	N	O
● Very large	30 to 100	0.77 to 1.00	P	Q	R

First, it is very relevant that comparisons take into account the reliability of the information from the geographical units (Goldstein and Spiegelhalter, 1996). Traditional estimations of area averages are unreliable if the number of individuals in the areas is small. In contrast, in multilevel regression the area averages are reliability-weighted estimates (Hox, 2002) and, therefore, more appropriated for comparisons.

Second, a correct interpretation of the plots and maps needs that the size of the geographical differences are evaluated by suitable measures such as the VPC and the AUC.

3.3. Quantifying the size of the geographical differences

Traditional geographical comparisons are based on differences between average values across areas (e.g., average BMI, percentage of individuals who are overweight/obese, simple measures of geographical variance). In these comparisons, the criteria for quantifying the size of the geographical differences are not clearly stated. For instance, when can we say that there is “substantial” geographical variation? When are the geographical differences small or negligible? The criteria of “statistical significance” is insufficient as very small differences between averages may nonetheless prove statistically significant if the sample is large.

In the MAIHDA framework we propose, we do not consider the differences between areas and differences between individuals as if they were two separate and unrelated phenomena of interest. Rather, we adopt a multilevel perspective that disentangles the absolute individual variance into a between area and a within area component. From this perspective, we evaluate geographical differences by quantifying the share of the total variation in individual outcomes that operates at the area level. In the case of binary individual health outcomes, we additionally use the AUC as it provides analogous information, but expressed in terms of discriminatory accuracy.

Currently there is no official guidance for assessing the magnitude of the VPC or the AUC in the context of studying area differences in individual health but a practical proposal is laid out in Table 1. This Table 1 also shows the corresponding AUC values according to the simulated relationship presented in Fig. 2. The proposed values are based on the authors' own experience but further discussion is needed to arrive at a standard classification. Furthermore, different standards may ultimately be required and developed for different health outcomes in different contexts and at different points in time. In any case, we recommend authors always report and discuss the exact values of the VPC/AUC and the exact difference from the benchmark.

3.4. Interpretation of the results

Using the concepts explained above; imagine we aim to evaluate a specific public health indicator such as BMI or overweight/obese status and we conduct a geographical analysis of the municipalities in a

county. For this geographical analysis, we need at least two types of information. First, we need to know to what degree the average health indicator (e.g., the average BMI or the prevalence of overweight/obese individuals) has reached a predetermined target level. To be precise, has the target been insufficiently, closely, or fully reached? Second, we need information on the *size of the observed geographical differences* as assessed by the VPC or the AUC. These two types of information need to be combined for the evaluation. Table 1 illustrates these ideas by means of a simple framework with 18 scenarios that can be used to orient the interpretation of an analysis.

In the ideal scenario (scenario C) the desired target level has been fully achieved in the county, and the municipality differences are effectively absent. The conclusion would be that all municipalities have performed similarly well.

In the worst scenario (scenario A) the desired target level has not been achieved in the county, and the municipality differences are again absent. The conclusion would be that all municipalities have performed similarly badly.

Observe that in both scenarios A and C interventions targeted to specific areas are not justifiable. Rather any intervention should be universal (i.e., directed to the whole county). In scenario C, the reason for the intervention would be to maintain the desirable level of health in all the municipalities. In the scenario A, the reason would be to achieve the desirable level of health in all the municipalities.

The interpretation of the scenarios in the lowest corners of the table (P and R) is conditioned by the very large sized of the geographical differences. For example, in scenario R some municipalities may have not achieved the target level even if the county as a whole has done so. In contrast, in scenario P, some municipalities may have achieved the target level even if the county as a whole has not achieved it. In these scenarios (P and R) targeted municipality interventions are justified.

The framework we propose fits well with the concept of *proportionate universalism* discussed by Marmot within resource allocation in public health (Marmot and Bell, 2012). That is, health actions must be

universal, not targeted, but with a scale and intensity that is proportionate to the level of disadvantage. The multilevel analytical approach we apply in this study can be used to inform decisions regarding the appropriate scale and intensity for a given geographic context.

4. Two real examples: geographical differences in BMI and obesity

4.1. Differences in BMI

At the time of the survey, the average BMI in Skåne was 25.4 kg/m² and ranged from 25.1 kg/m² in the municipality number 1 to 26.7 kg/m² in the municipality number 33. Fig. 3 shows the differences in average BMI across the 33 Skåne municipalities. The average BMI is very close to the target level of 25.0 kg/m² in the county as a whole and in most of its municipalities. The Fig. 3 shows the differences in average BMI between the 33 Skåne municipalities.

The multilevel linear regression showed a low VPC value of 0.9%, which according to the framework propose above (Table 1) suggests that meaningful geographical differences are absent. In addition, the county average value is just above the target. Therefore, the results are closest to scenario B (Table 1) and we can conclude that all municipalities have performed homogeneously well. The municipality context is an important public health arena for intervention and our results suggest that all municipalities must work to keep the existing level of homogeneity and prevent future disparities.

To support this idea, Fig. 4 plots the distribution of *individual* BMI in each municipality in Skåne. Each point represents a different individual in the data. The horizontal line represents the target BMI value of 25.0 kg/m². In agreement with the low reported VPC, Fig. 4 shows that there is considerable overlap between municipalities' individual BMI's distributions. This means that if we based our public health efforts on differences in area average values, many people with a high BMI in the "best" municipality (i.e., municipality number 1) would be overlooked if we focused only on the municipality with the highest average BMI

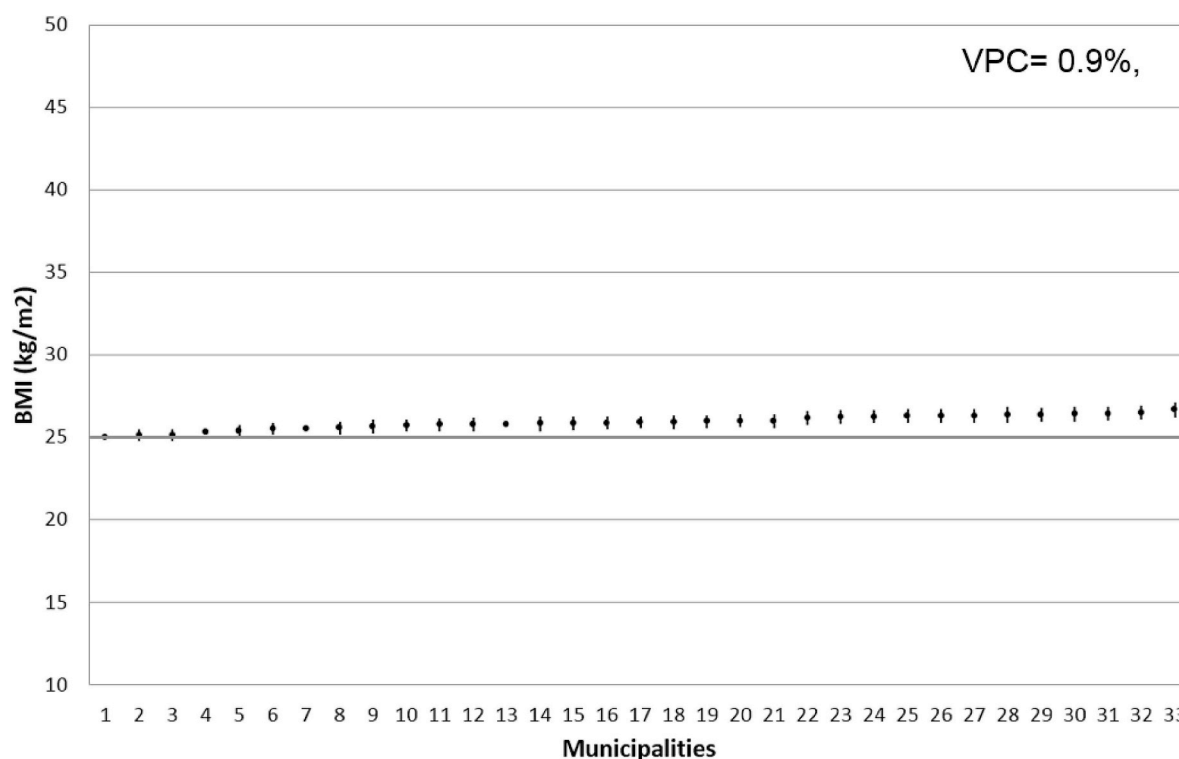


Fig. 3. Differences in average body mass index (BMI) (small black circles) with 95% confidence intervals (small vertical lines crossing the circles) between the 33 Skåne municipalities based on the Public Health Survey for Skåne 2008 with 27,222 participants aged between 18 and 80 years. The variance partition coefficient (VPC) indicates that 0.9% of the total individual BMI variation is at the municipal level.

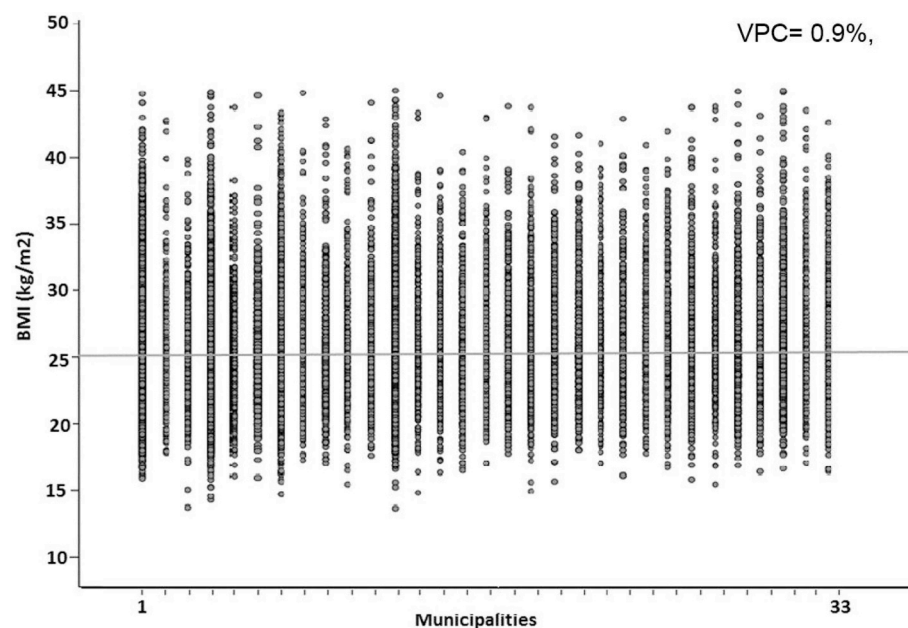


Fig. 4. Differences in the occurrence of obesity (i.e., body mass index (BMI) equal or higher than 25 kg/m^2) among the 33 Skåne municipalities based on the Public Health Survey for Skåne 2008 with 27,222 participants aged between 18 and 80 years. The circles represent the specific individual values. The variance partition coefficient (VPC) indicates that 0.9% of the total individual BMI variation at the municipal level.

values (i.e., municipality number 33). According to the model, 38% of the individuals in municipality number 1 are predicted to have a higher BMI than the average BMI in municipality number 33. Analogously, 38% of the individuals in municipality number 33 have a BMI lower than the average BMI in municipality number 1. These statistics make clear that the “bad” municipality (number 33) should not be singled out and as a result unfairly labelled. In fact, the analysis of these data suggests that public health efforts to influence citizens' BMI need to be universal rather than targeted to only the highest BMI municipalities of Skåne.

4.2. Differences in obesity

When the health indicator is a binary individual outcome, the area level averages are proportions (e.g., prevalence, absolute risk). Therefore, in this example we analyse overweight/obese prevalence rather than average BMI to compare municipalities. Fig. 5 shows that the prevalence of overweight/obesity in Skåne as a whole was 51% and ranged from 44% in municipality number 1 to 63% in municipality number 33. Of the 33 municipalities, nine show an overweight/obesity prevalence below 50%.

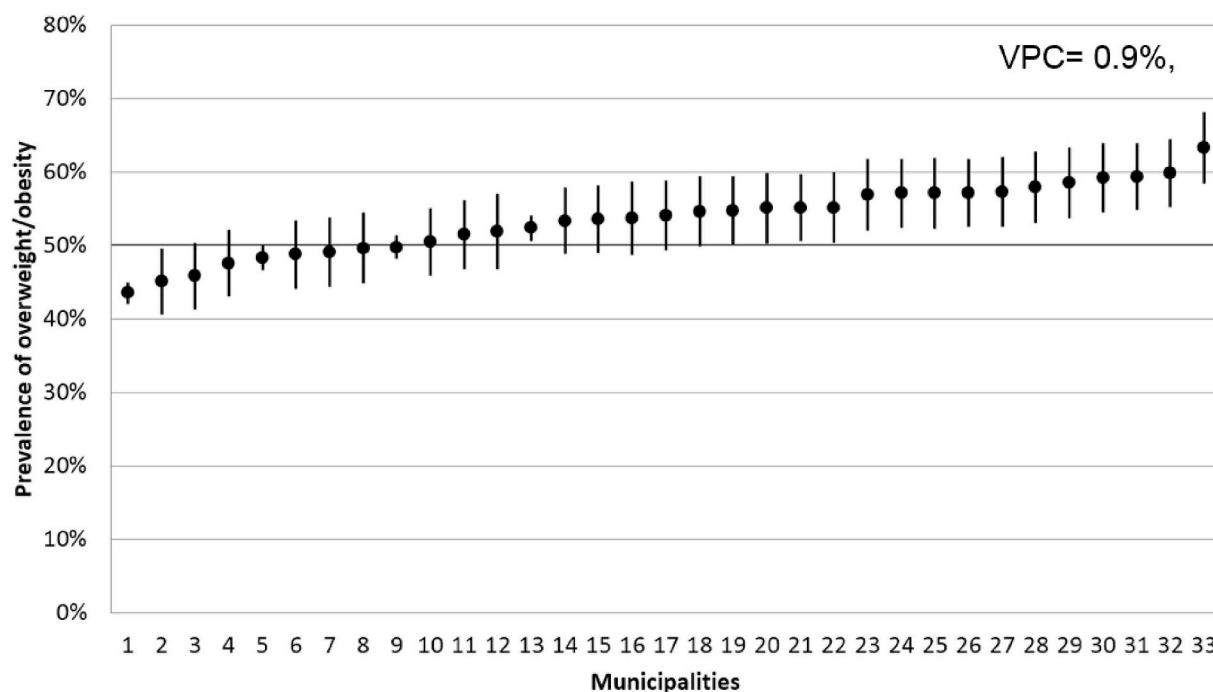


Fig. 5. Differences in the occurrence of obesity (i.e., body mass index (BMI) equal or higher than 25 kg/m^2) among the 33 Skåne municipalities based on the Public Health Survey for Skåne 2008 with 27,222 participants aged between 18 and 80 years. The black circles represent the predicted municipality prevalence values with 95% confidence intervals (small vertical lines crossing the circles). The variance partition coefficient (VPC) indicates that 0.9% of the total individual BMI variation at the municipal level.

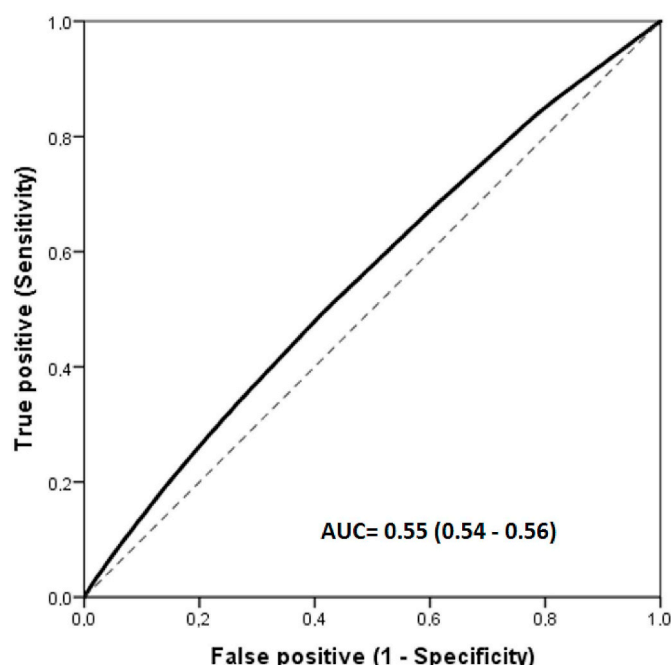


Fig. 6. Area under the receiver operating characteristic curve (AUC) with (95% confidence interval) constructed by plotting the true positive fraction (TPF) (i.e., sensitivity) against the false positive fraction (FPF) (i.e., $1 - \text{specificity}$) for different binary classification thresholds of the predicted probabilities based on the municipality random effects from the multilevel logistic regression analysis. The AUC measures the accuracy of knowing where the individual resides (i.e., the municipalities) for discriminating individuals (i.e., between overweight/obese individuals and normal weight individuals). The AUC takes a value between 0.5 and 1.0 where 0.5 is a complete lack of discrimination and where 1.0 is perfect discrimination.

The results of the multilevel logistic regression showed a VPC of 0.9%, which, as expected, was similar to the VPC in the multilevel linear regression. Thus, according to our framework (Table 1), a negligible component of the total individual variation in the underlying latent propensity of being overweight/obese was at the municipal level.

The result of the AUC analysis in our example is shown in Fig. 6. The AUC is equal to 0.55. This value is the expected value given the relationship between the VPC and the AUC (Fig. 1) and again is a very low value according to our framework (Table 1). The conclusion is the same as when interpreting the VPC. That is, information on municipality differences in overweight/obese prevalence provides extremely inaccurate information for the purpose of identifying and targeting overweight/obese individuals.

5. Discussion

The present study crystallizes ideas discussed in a body of previous publications revisiting traditional epidemiological methods used for geographical comparisons in epidemiology and public health (Merlo et al., 2001a,b; Merlo et al., 2004; Ohlsson et al., 2011; Merlo et al., 2012; Merlo et al., 2016). These ideas, however, also apply to comparisons between healthcare facilities such as hospitals (Merlo et al., 2001a,b; Ghith et al., 2016) and healthcare centres (Hjerpe et al., 2011; Ohlsson and Merlo, 2007) and even when the analysis of health differences is based on other grouping criteria such as ethnicity (Beckman et al., 2004; Mulinari et al., 2015) or socioeconomic status (Axelsson Fisk and Merlo, 2018a) or the combination of such criteria (Wemrell et al., 2017; Hernandez-Yumar et al., 2018).

Fundamentally, the MAIHDA framework we promote stresses the systematic analysis of not only differences between group averages but also of the individual heterogeneity around those averages. By doing so

we obtain a better understanding of the distribution and determinants of individual health and disease risk in the population. This analytical approach converges with the current movement of precision (i.e., individualized, personalized, stratified) medicine and public health, and their efforts toward understanding not only differences in average health outcomes between population groups, but also individual heterogeneity around such averages. However, as discussed elsewhere (Khouri and Galea, 2017; Chowkwanyun et al., 2018), the fundamental question is if the move towards “precision” public health offers the opportunity for a reconceptualised, empowered public health enterprise — or might it represent an abandonment of the traditional mission of enhancing population well-being? From this perspective, a fundamental conceptual distinction exists between the MAIHDA and the individualized medicine approaches: *rather than focusing on individual biomedical susceptibilities, MAIHDA tries to identify the components of individual heterogeneity in health that operate at different contextual, spatial and organisational levels and which evolve across the life-course*. In this way, the MAIHDA approach aims to harmonize the study of group averages with the study of individual heterogeneity around these averages. Our approach rests on the fundamental notion that individual and population health are not dislocated phenomena of interest. Rather, it considers the existence of a continuous distribution of individual outcome heterogeneity that can be articulated at different levels of analysis (Merlo, 2014).

Applying a simple analytic framework based on both the overall average (i.e., county BMI average or prevalence of overweight/obese) and the VPC/AUC, we have provided an efficient methodology for evaluating geographical differences in public health reports. In our empirical examples, we showed how traditional analyses based only on area averages would have identified and labelled some municipalities as worse (e.g., municipality number 33) than others (e.g., municipality number 1). However, the multilevel analysis shows that such average differences, while present, are trivial once one acknowledges the substantial variation in individual health outcomes within areas.

Nevertheless, it could be reasonably argued that the absolute size of the geographical variance provides relevant information by itself (Subramanian and O'malley, 2010). We think it is necessary to interpret both the area variance and the area VPC simultaneously. For example we plot the BMI and obesity means in Figs. 3 and 5 which are graphical depictions of the area variance. It is not a matter of choosing between the area variance or the area VPC rather to consider both simultaneously. We learn more from the data by looking at both statistics using each to interpret the other. However, we think we should put more emphasis on the VPC as it may have higher public health relevance. We believe our study illustrates this situation.

A between-area variance component that might be assumed large in absolute terms may nonetheless prove to be small in relation to the total absolute individual-level variation (i.e., the sum of the between-area and within-area variance components). That is, the VPC and AUC may both be low. Therefore, the exclusive quantification of geographical differences by the area variance neglects the information provided by the VPC/AUC which, we think, are of high public health relevance. Our study illustrates this situation.

The area variance is a measure that summarizes differences between area averages and its interpretation is analogous to other measures of differences between averages such as the odds ratio. In 2004 Pepe et al. (Pepe et al., 2004) demonstrated that an exposure we believe is strongly associated with the outcome (e.g., odds ratio = 10) is not necessarily effective for classifying persons according to that outcome, as we often assume. For this assumption to be true, the odds ratio must be of a much higher magnitude than that we use to consider as high. Making an analogy, a geographical variance that we assume to be large may have a low VPC and a low AUC.

If the VPC/AUC is low, someone may ask why focus on *those* areas? This is not an erroneous question, see for instance (Boyle and Willms, 1999). However, it would be erroneous to argue that overall places

do not matter because one specific area definition shows a low VPC for a specific outcome. The proper reaction would be to start searching for better geographical definitions of the context that influence the outcome of interest or to even combine geographical and social information to better define contexts. We need more a priori theory to define geographical and other contexts in relation to specific outcomes.

As far we know, the question on how we can evaluate the size of the area variance has not been satisfactorily answered (Ibanez et al., 2009; Diehr et al., 1990). Therefore, a fundamental contribution of our paper is to show that measures of components of variance and discriminatory accuracy assist us when quantifying the size of the area variance. Using this new knowledge we may reconsider our previous assumptions for what is a large or a small geographical variation.

However, it is worth stressing that our finding of small place effects on BMI concerns only the municipalities in Skåne. In addition, while certain administrative geography might not prove relevant for one health outcome, it may well prove important for others. Absent or small municipality differences do not necessarily indicate a lack of geographical differences or of “place effects” more generally (Merlo et al., 2016; Boyle and Willms, 1999). For example, our finding may reflect that the administrative municipality boundaries we used do not truly capture the relevant contexts that influence individual BMI. Other administrative scales or different context definitions may evidence higher VPCs and AUCs. From this perspective it is recommended to use theory and prior research to guide methodological decisions (e.g., choice of scale), and to generate hypotheses regarding the spatial processes that may cause geographic differences (e.g., contagion, shared built environments, etc.) and that may drive the relative contributions of individual and geographic differences for a given health outcome. It is likely that BMI is simultaneously influenced by many contexts including the household, the school, and the work place and the influences of these contexts may further interact with one another. From this perspective, the small municipality differences are perhaps not surprising.

Our study focuses on the effects of the general context defined by municipality boundaries on individual BMI/overweight-obesity status and a key measure is the VPC. Therefore, it is relevant to consider the possibility of measurement error in the outcome variable at the individual level. If measurement error exists, the estimated individual level variance in two-level individual-within-area models will be biased upwards leading the estimated area level VPC to be biased downwards and for the area effects to therefore appear less important than they truly are (Subramanian and O'malley, 2010). In our case, BMI information is based on self-reported height and weight, which appears to have an acceptable validity and reliability (Nyholm et al., 2007) and so any downwards-bias in the estimation of the municipality VPC should not invalidate the conclusions of our study. However, we encourage researchers to always strive to obtain outcomes (as well as exposures) with a high validity and reliability to obtain correct estimations of variance components.

Our study illustrates that by neglecting individual variation around averages, traditional epidemiological analyses could ultimately lead to ‘wrong’ public health decisions in particular if resources were completely targeted to the municipalities with the highest BMIs. However, in practice, this extreme targeting is infrequent and proportionate universalism is now a accepted alternative for resource allocation in public health (Carey et al., 2015). As Sir Michael Marmot advocated (Marmot and Bell, 2012), health actions must be universal, not targeted, but with a scale and intensity that is proportionate to the level of disadvantage. The MAIHDA approach can be used to inform decisions regarding the degree to which public health interventions need to be universally targeted.

6. Conclusion

Our study aims to provide a clear, easy-to-follow, applied demonstration of how multilevel analysis and the MAIHDA approach

including interpretation of the VPC/AUC can aid the formulation of geographical health policy. Multilevel modelling can also provide improved point estimates and confidence intervals for (adjusted) and offers methodological advantages when ranking areas, for example, when constructing league tables (Leckie and Goldstein, 2011). It may also be useful as the basis to create funnel plots and risk maps. However, here too, this information should always be accompanied with measures of discriminatory accuracy such as the VPC or the AUC (Merlo et al., 2016). The multilevel analytical framework we propose avoids the “tyranny of the averages” and by providing more nuanced information, it reduces the risk of inaccurate labelling of areas, which may itself have negative public health influences. Finally, if the VPC associated with a specific geographical context is low, researchers might use theory to identify alternative geographical contexts. It is also possible to consider non-geographical contexts defined by different combinations of social and economic dimensions or contexts defined by the combination of geographical and socioeconomic categorizations. For example, an emerging research field is applying MAIHDA combined with intersectionality theory (Axelsson Fisk et al., 2018; Hernandez-Yumar et al., 2018; Evans and Erickson, 2018; Evans et al., 2018) in order to increase our understanding of population heterogeneity. See elsewhere for an introduction to this complementary approach (Merlo, 2014; Merlo, 2018b).

Applying MAIHDA to analyse geographical differences in health presents a number of advantages over traditional analytical approaches. From the perspective of proportionate universalism, MAIHDA can help us to identify the appropriate scale and intensity of geographical public health interventions and, thereby, help us to decide whether targeted or universal interventions should be pursued.

Funding

The Swedish Research Council (VR) (<https://www.vr.se/english.html>) supported this work through the project “Multilevel Analyses of Individual Heterogeneity: innovative concepts and methodological approaches in Public Health and Social Epidemiology”: (#2017-01321, JM).

Declarations of interest

None.

Acknowledgements

We would like to express our gratitude to Prof Karien Stronks (Department of Public Health; Academic Medical Centre, University of Amsterdam), Carl Lundgren, Åsa Ljungvall (Swedish Agency for Health and Care Services Analysis), Sten Axelsson Fisk and Maria Wemrell (Unit for Social Epidemiology, Faculty of Medicine, Lund University) for reading and commenting on earlier versions of this work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.healthplace.2019.102145>.

References

- Axelsson Fisk, S., Merlo, J., 2017. Absolute rather than relative income is a better socioeconomic predictor of chronic obstructive pulmonary disease in Swedish adults. *Int. J. Equity Health* 16, 70.
- Axelsson Fisk, S., Mulinari, S., Wemrell, M., Leckie, G., Perez-Vicente, R., Merlo, J., 2018. Chronic Obstructive Pulmonary Disease in Sweden: an intersectional multilevel analysis of individual heterogeneity and discriminatory accuracy. *Epub* 2018/06/02. *Soc. Sci. Med. (Population Health)* 334–346. <https://doi.org/10.1016/j.ssmph.2018.03.005>. PubMed PMID: 29854918. PubMed Central PMCID: PMC5976844.
- Beckman, A., Merlo, J., Lynch, J.W., Gerdtham, U.G., Lindstrom, M., Lithman, T., 2004. Country of birth, socioeconomic position, and healthcare expenditure: a multilevel

- analysis of Malmö, Sweden. *J. Epidemiol. Community Health* 58, 145–149.
- Boyle, M.H., Willms, J.D., 1999. Place effects for areas defined by administrative boundaries. *Am. J. Epidemiol.* 149, 577–585.
- Carey, G., Crammond, B., De Leeuw, E., 2015. Towards health equity: a framework for the application of proportionate universalism. *Int. J. Equity Health* 14, 81.
- Centre for multilevel modelling u. O. B., <http://www.bristol.ac.uk/cmm/learning/online-course-LEMMMA> (Learning Environment for Multilevel Methodology and Applications) online multilevel modelling course [Online]. [Accessed 2018].
- Chowkwanyun, M., Bayer, R., Galea, S., 2018. "Precision" public health - between novelty and hype. *N. Engl. J. Med.* 379, 1398–1400.
- Diehr, P., Cain, K., Connell, F., Volinn, E., 1990. What is too much variation? The null hypothesis in small-area analysis. *Health Serv. Res.* 24, 741–771.
- Duncan, C., Jones, K., Moon, G., 1998. Context, composition and heterogeneity: using multilevel models in health research. *Soc. Sci. Med.* 46, 97–117.
- Durkheim, E., 1964. *The Rules of Sociological Method*. Free Press of Glencoe, New York.
- Evans, C., Erickson, N., 2018. Intersectionality and depression in adolescence and early adulthood: a MAIHDA analysis of the national longitudinal study of adolescent to adult health, 1995–2008. *Soc. Sci. Med.* 220, 1–11.
- Evans, C.R., Williams, D.R., Onnela, J.P., Subramanian, S.V., 2018. A multilevel approach to modeling health inequalities at the intersection of multiple social identities. *Epub* 2017/12/05. *Soc. Sci. Med.* 203, 64–73. <https://doi.org/10.1016/j.socscimed.2017.11.011>. PubMed PMID: 29199054.
- Ghith, N., Wagner, P., Frolich, A., Merlo, J., 2016. Short term survival after admission for heart failure in Sweden: applying multilevel analyses of discriminatory accuracy to evaluate institutional performance. *PLoS One* 11, e0148187.
- Goldstein, H., Spiegelhalter, D., 1996. League tables and their limitations: statistical issues in comparisons of institutional performance. *J. R. Stat. Soc. Ser. A* 159, 385–443.
- Goldstein, H., Browne, W., Rasbash, J., 2002. Partitioning variation in generalised linear multilevel models. *Understand. Stat.* 1, 223–232.
- Hernandez-Yumar, A., Wemrell, M., Abasolo Alesson, I., Gonzalez Lopez-Valcarcel, B., Leckie, G., Merlo, J., 2018. Socioeconomic differences in body mass index in Spain: an intersectional multilevel analysis of individual heterogeneity and discriminatory accuracy. *PLoS One* 13, e0208624.
- Hjerpe, P., Ohlsson, H., Lindblad, U., Bostrom, K.B., Merlo, J., 2010. Understanding adherence to therapeutic guidelines: a multilevel analysis of statin prescription in the Skaraborg Primary Care Database. *Epub* 2010/12/31. *Eur. J. Clin. Pharmacol.* 67 (4), 415–423. <https://doi.org/10.1007/s00228-010-0973-4>. PubMed PMID: 21190018.
- Hox, J.J., 2002. *Multilevel Analysis Techniques and Applications*. Lawrence Erlbaum Associates, Mahwah, N.J.
- Ibanez, B., Librero, J., Bernal-Delgado, E., Peiro, S., Lopez-Valcarcel, B.G., Martinez, N., Aizpuru, F., 2009. Is there much variation in variation? Revisiting statistics of small area variation in health services research. *BMC Health Serv. Res.* 9, 60.
- Khoury, M.J., Galea, S., 2017. Evaluating precision medicine's ability to improve population health-reply. *J. Am. Med. Assoc.* 317, 441.
- Leckie, G., Goldstein, H., 2011. Understanding uncertainty in school league tables. *Fisc. Stud.* 32 (2), 207–224. <https://doi.org/10.1111/j.1475-5890.2011.00133>.
- Marmot, M., Bell, R., 2012. Fair society, healthy lives. *Publ. Health* 126 (Suppl. 1), S4–S10.
- Merlo, J., 2003. Multilevel analytical approaches in social epidemiology: measures of health variation compared with traditional measures of association. *J. Epidemiol. Community Health* 57, 550–552.
- Merlo, J., 2014. Invited commentary: multilevel analysis of individual heterogeneity—a fundamental critique of the current probabilistic risk factor epidemiology. *Am. J. Epidemiol.* 180, 208–212 discussion 213–4.
- Merlo, J., 2018a. Multilevel analysis of individual heterogeneity and discriminatory accuracy (MAIHDA) within an intersectional framework. *Soc. Sci. Med.* 203, 74–80. *Epub* 2018/01/07. <https://doi.org/10.1016/j.socscimed.2017.12.026> PubMed PMID: 29305018.
- Merlo, J., 2018b. Multilevel analysis of individual heterogeneity and discriminatory accuracy (MAIHDA) within an intersectional framework. *Soc. Sci. Med.* 203, 74–80.
- Merlo, J., Mulinari, S., 2015. Measures of discriminatory accuracy and categorizations in public health: a response to Allan Krasnik's editorial. *Eur. J. Public Health* 25, 910.
- Merlo, J., Broms, K., Östergren, P.O., Hagberg, O., Nordlund, A., Lithman, T., 2001a. Multinivåanalys av regionala skillnader i överlevnad efter hjärtsvikt: landstingens egenskaper har foga inflytande på patientens prognos. *Lakartidningen* 98, 4838–4844.
- Merlo, J., Östergren, P.O., Broms, K., Björck-Linne, A., Liedholm, H., 2001b. Survival after initial hospitalisation for heart failure: a multilevel analysis of patients in Swedish acute care hospitals. *J. Epidemiol. Community Health* 55, 323–329.
- Merlo, J., Asplund, K., Lynch, J., Rastam, L., Dobson, A., 2004. Population effects on individual systolic blood pressure: a multilevel analysis of the World Health Organization MONICA Project. *Am. J. Epidemiol.* 159, 1168–1179.
- Merlo, J., Chaix, B., Yang, M., Lynch, J., Rastam, L., 2005a. A brief conceptual tutorial of multilevel analysis in social epidemiology: linking the statistical concept of clustering to the idea of contextual phenomenon. *J. Epidemiol. Community Health* 59, 443–449.
- Merlo, J., Chaix, B., Yang, M., Lynch, J., Rastam, L., 2005b. A brief conceptual tutorial on multilevel analysis in social epidemiology: interpreting neighbourhood differences and the effect of neighbourhood characteristics on individual health. *J. Epidemiol. Community Health* 59, 1022–1028.
- Merlo, J., Yang, M., Chaix, B., Lynch, J., Rastam, L., 2005c. A brief conceptual tutorial on multilevel analysis in social epidemiology: investigating contextual phenomena in different groups of people. *J. Epidemiol. Community Health* 59, 729–736.
- Merlo, J., Chaix, B., Ohlsson, H., Beckman, A., Johnell, K., Hjerpe, P., Rastam, L., Larsen, K., 2006. A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *J. Epidemiol. Community Health* 60, 290–297.
- Merlo, J., Ohlsson, H., Lynch, K.F., Chaix, B., Subramanian, S.V., 2009. Individual and collective bodies: using measures of variance and association in contextual epidemiology. *J. Epidemiol. Community Health* 63, 1043–1048.
- Merlo, J., Viciana-Fernandez, F.J., Ramiro-Farinas, D., Research Group Of Longitudinal Database Of Andalusian, P., 2012. Bringing the individual back to small-area variation studies: a multilevel analysis of all-cause mortality in Andalusia, Spain. *Soc. Sci. Med.* 75, 1477–1487.
- Merlo, J., Wagner, P., Ghith, N., Leckie, G., 2016. An original stepwise multilevel logistic regression analysis of discriminatory accuracy: the case of neighbourhoods and health. *PLoS One* 11, e0153778.
- Merlo, J., Mulinari, S., Wemrell, M., Subramanian, S.V., Hedblad, B., 2017. The tyranny of the averages and the indiscriminate use of risk factors in Public Health: the case of coronary heart disease. *SSM Popul. Health* 3, 684–698.
- Mulinari, S., Juarez, S.P., Wagner, P., Merlo, J., 2015. Does maternal country of birth matter for understanding offspring's birthweight? A multilevel analysis of individual heterogeneity in Sweden. *PLoS One* 10, e0129362.
- Nyholm, M., Gullberg, B., Merlo, J., Lundqvist-Persson, C., Rastam, L., Lindblad, U., 2007. The validity of obesity based on self-reported weight and height: implications for population studies. *Obesity (Silver Spring)* 15, 197–208.
- Ohlsson, H., Merlo, J., 2007. Understanding the effects of a decentralized budget on physicians compliance with guidelines for statin prescription; a multilevel methodological approach. *BMC Health Serv. Res.* 7, 68.
- Ohlsson, H., Librero, J., Sundquist, J., Sundquist, K., Merlo, J., 2011. Performance evaluations and league tables: do they capture variation between organizational units? An analysis of 5 Swedish pharmacological performance indicators. *Med. Care* 49, 327–331.
- Pepe, M.S., Janes, H., Longton, G., Leisenring, W., Newcomb, P., 2004. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am. J. Epidemiol.* 159, 882–890.
- Rose, G.A., 1992. *The Strategy of Preventive Medicine*. Oxford University Press, Oxford England ; New York.
- Rosling, H., Rosling, O., Rönnlund, A.R., 2018. *Factfulness : Ten Reasons We're Wrong about the World –AndWhy Things Are Better than You Think*. Flatiron Books, New York.
- Rosvall, M., Grahn, M., Modén, B., Author, J., 2009. *Hälsöförhållanden i Skåne Folkhälsoenkät Skåne 2008. Socialmedicinska enheten. Region Skåne*. http://utveckling.skane.se/siteassets/publikationer_dokument/folkhalsoenkat_skane_2008.pdf?id=2535.
- Snijders, T.A.B., Bosker, R.J., 1999. *Multilevel Analysis: an Introduction to Basic and Advanced Multilevel Modeling*. SAGE Publications, Thousand Oaks, California.
- Subramanian, S.V., O'malley, A.J., 2010. Modeling neighborhood effects: the futility of comparing mixed and marginal approaches. *Epidemiology* 21, 475–478 discussion 479–81.
- Subramanian, S.V., Jones, K., Duncan, C., 2003. Multilevel methods for public health research. In: Kawachi, I., Berkman, L.F. (Eds.), *Neighborhoods and Health*. Oxford University Press, New York, NY.
- The Commission For Health On Equal Terms, 2017. Next step towards increasing health on equal terms. Proposals for long-term work for a good health on equal terms. Final report by the Commission for Health on Equal Terms (Nästa steg på vägen mot en mer jämlik hälsa. Förslag för ett långsiktigt arbete för en god och jämlik hälsa - slutbetänkande av Kommissionen för jämlik hälsa). Statens offentliga utredningar.
- Wemrell, M., Mulinari, S., Merlo, J., 2017. Intersectionality and risk for ischemic heart disease in Sweden: categorical and anti-categorical approaches. *Soc. Sci. Med.* 177, 213–222.